



University of Groningen

Prediction intervals for missing figures in migration tables

Jong, Paul Martin de

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2008

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Jong, P. M. D. (2008). Prediction intervals for missing figures in migration tables. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

SAMENVATTING

Voor de specificatie van bepaalde bevolkingsmodellen zijn jaarlijkse aantallen binnenlandse migraties naar herkomst, bestemming en leeftijd nodig terwijl de meeste nationale statistische instituten slechts randtotalen van zulke 3-dimensionale tabellen publiceren. Deze discrepantie was de aanleiding tot dit onderzoek naar het reconstrueren van ontbrekende cijfers in frequentietabellen.

In hoofdstuk 1 wordt dit probleemgebied verkend. Er wordt gesteld dat men zich meestal beperkt tot het maken van puntschattingen. Zulke ramingen hebben echter weinig wetenschappelijke waarde. Daarom proberen wij intervallen te vervaardigen waarvan in theorie ongeveer 95% de werkelijke maar onbekende uitkomst zullen bevatten. Onze theorie is hoofdzakelijk gebaseerd op Poisson aannames en log-lineaire modellen.

Hoofdstuk 2 is geconcentreerd rond de gedachte dat het wiskundig gezien zeer geriefelijk is als de voorwaardelijke verdeling van de onbekende cijfers, gegeven de data, onafhankelijk is van de onbekende parameters in het model. Daarom worden statistische modellen gebouwd die aan deze eis voldoen. Zulke modellen stellen ons in staat om "exacte" voorspellingsintervallen en goede benaderingen hiervoor te produceren.

Vaak is de situatie echter ingewikkelder doordat er van nature een model gegeven is dat niet aan de gestelde eis voldoet. Voor zulke situaties wordt in hoofdstuk 3 een algemene aanpak ontwikkeld. Onze voorstellen zijn gecontroleerd voor enkele speciale gevallen en kunnen ook beschouwd worden als een veralgemening van sommige resultaten in hoofdstuk 2. Toch zijn we niet geheel tevreden omdat de benodigde benaderingen niet onderbouwd zijn door algemene asymptotische theorie.

Een voorwaarde om tot resultaten te komen is dat het aantal onbekende parameters in het model niet het aantal beschikbare cijfers mag overschrijden. Dit kan leiden tot modellen die te ver bezijden de waarheid zijn. Daarom worden er in deel II van dit proefschrift enkele concrete toepassingen bestudeerd van de in deel I ontwikkelde theorie. Er wordt gebruik gemaakt van volledige migratietabellen omdat deze ons in staat stellen de uit onvolledige informatie geconstrueerde intervallen te

confronteren met de werkelijke uitkomsten.

In hoofdstuk 4 wordt een 3-dimensionale tabel gereconstrueerd op basis van de drie 2-dimensionale tabellen van randtotalen. Het blijkt dat 94% van onze intervallen de werkelijke uitkomst bevatten. Het onderliggende model en de bijbehorende theorie lijken dus betrouwbaar. Gaan we echter reconstrueren op basis van veel minder gegevens, dan zijn extra modelaannames nodig en blijken veel meer dan 6% van de intervallen fout.

In hoofdstuk 5 wordt de interregionale migratietabel van een bepaald jaar biproportioneel aangepast aan de kolom- en rijtotalen van de tabel voor het daaropvolgende jaar. Dit levert bevredigende resultaten op voor dat nieuwe jaar. Indien we, in plaats van over de aantallen in- en uitmigraties, slechts over de migratiesaldi voor het nieuwe jaar beschikken, dan wordt "updaten" twijfelachtig.

Onze theorie heeft ook toepassingen buiten het gebied van migratie. In hoofdstuk 6 gaat het om het aantal schepen en hun totale laadvermogen dat jaarlijks een bepaald punt passeert. Omdat het te kostbaar is om deze cijfers te verkrijgen door gedurende het hele jaar exact te tellen worden ze geschat op basis van steekproefgegevens. Door het construeren van voorspellingsintervallen voor de werkelijke jaarcijfers wordt de onzekerheid in deze ramingen aangegeven.

14644
1985